What's "This" You Say? The Use of Local References on Distant Displays

Patti Bao, Darren Gergle
Center for Technology and Social Behavior
Northwestern University
2240 Campus Drive, Evanston, IL 60208 USA
{pattibao, dgergle}@northwestern.edu

ABSTRACT

This study explores how the design of visual display configurations relates to linguistic expressions. Twenty-five participants performed a series of object identification and narrative Description tasks on either a large wall-sized or small desktop display. Results revealed that during the Description tasks, large display users produced significantly greater rates of local deictic references than small display users, but in the identification tasks, the rates were similar for both large and small display users. Implications for the design of interactive technologies are discussed.

Author Keywords
Large display, field of view, language, discourse

ACM Classification Keywords
H5.1. **[Information Interfaces and Presentation (HCI)]**: Multimedia Information Systems.

INTRODUCTION

The way we speak and what we comprehend is intimately tied to the visual context in which we communicate. Without a thorough understanding of how language and visual information interrelate, we run the risk of developing technologies that fail to support natural interactions. For example, an automated conversational agent needs to generate speech and actions in line with natural human behavior. When this does not transpire, studies have shown that people adapt their communication patterns (e.g., by using hyper-articulated speech) in ways that lead to additional difficulties for computing systems [9].

In this paper we examine how display size, controlling for field of view, can significantly influence language use and ultimately affect our interactions with visual environments. We build on previous findings by Tan and colleagues [12] that suggest larger displays, independent of field of view, can invoke an egocentric perspective that influences a range of behavioral outcomes such as spatial knowledge and memory. If an egocentric perspective is taken then we should also expect linguistic adaptations that demonstrate this shift in perspective. Despite this seemingly natural connection between cognitive perspective and language use, surprisingly little work has been done to examine how particular display configurations relate to the pattern of observable linguistic expressions.

We report results that show how display size influences the distribution of particular referential forms used to describe the spatial attributes of an environment. We discuss how these findings have important implications for improving the design of a wide range of technologies including those that engage in naturalistic conversation with humans [2], or those in which adaptations are based upon an understanding of the current dialogue state [10]. Finally, we propose that linguistic analysis offers a useful approach for researching the cognitive affordances of different technological factors across visual environments [see also 8].

BACKGROUND

Previous research has demonstrated benefits of using large displays that range from improved individual performance in complex tasks [3] and spatial tasks [11, 13] to greater shared awareness [4] and more fluid social interaction [7] among collocated collaborators. More recently, Tan and colleagues [12] examined tasks that could be performed from either an egocentric perspective (e.g., a 1st-person view in which one imagines themselves rotating within an environment) or an exocentric perspective (e.g., a 3rd-person view in which one imagines objects rotating around one another in space). They found that large displays biased users into an egocentric mindset associated with spatial performance benefits, while small displays encouraged an exocentric perspective without the associated gains.

In the study presented in this paper we reason that if display size can influence the perspective that we bring to a visual task, then we should also see significant shifts in the way we *talk* about a given space. A handful of recent studies have demonstrated that the form of linguistic expressions

CHI 2009, April 4–9, 2009, Boston, MA, USA.
Copyright 2009 ACM 978-1-60558-246-7/09/04…$5.00.

changes depending on the visual context in which they are produced. For example, when people share visual space they shorten full noun phrase descriptions to deictic pronouns [6], and they shift the distribution of local and remote deixis (e.g., *this/here* vs. *that/there*) according to whether the speakers perceive themselves to be physically co-present [1, 8]. Byron & Stoia [1] provide evidence that spatial factors play a role during reference in virtual environments, showing that speakers tended to favor local deixis when they were closer to a given object and remote deixis when the hearer was closer to an object. Kramer and colleagues demonstrate a relationship between perceived presence and the use of local and remote deixis, and suggest that the interaction medium also has an effect [8].

In this study, we were particularly interested in whether display size alone, controlling for visual angle (i.e., a large image and a long viewing distance vs. a small image and a short viewing distance), could influence people's cognitive perspective and ultimately the referential forms they use to describe the objects and attributes of a virtual environment. We also expected characteristics of the communicative task along with spatial features of the environment to further influence deictic patterns (based on work by [1]).

METHOD

Participants

Twenty-five (13 female) students and staff members from a mid-sized Midwestern university participated in the study. Participants were native English speakers with normal or corrected-to-normal eyesight. Thirty-six percent were 18-21 years old, 32% were 22-25 years old, and 32% were 26-40 years old. Participants interacted with the massively multiplayer online game (MMOG) World of Warcraft (WoW). Forty-four percent of participants had never played a MMOG and 20% reported having played WoW before participating in the study. They were paid $10 an hour.

Procedure

The study design was a 2 (Display Size) × 2 (Task) × 2 (Object Distance) where Display Size was a between-participants factor, and Task and Object Distance were within-participants factors with presentation order counter-balanced across participants.

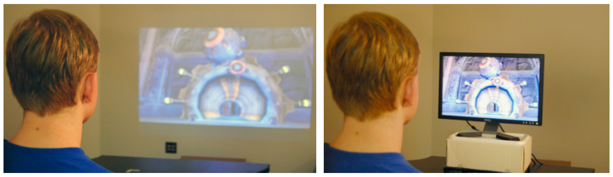

Figure 1. Large (left) and Small (right) display setups. (Note: Lights were off so the bezel was not visible.)

Upon entering the laboratory, participants were first asked to complete the ETS Card Test to establish a baseline measure of visual-spatial abilities [5]. They were then randomly assigned (balanced by gender) to either the Large or Small Display condition (see Figure 1). Within each display condition the participants performed repeated trials of two different task types.

In the **Description Task**, the participant produced a spoken narrative to accompany a pre-recorded two-minute video clip in which an avatar walked through the WoW environment in first-person perspective. In addition to this open-ended narrative, the experimenter paused the clip at preset intervals to prompt the participant about an entity on the screen. For example, when a knight character appeared in the scene, the experimenter asked: "What do you think about what he's wearing?" Care was taken in the formation of these prompts to avoid providing spatial information. The goal of this task was to elicit as much free-form spoken narrative about the environment as possible.

In the **Identification (ID) Task**, the participant viewed a 30 second video clip that led up to a static array of objects. For example, in Figure 2, a player walks into a storefront and stops just before the counter. The experimenter then asked the participant to locate the position of a pre-determined object, in this case using the prompt, "The object is a set of five similar things, each half-full of liquid." The participant was then given time to locate the object and describe its position. The goal of this task was to force the production of referential expressions rich in spatial information (e.g., "that one on the left" or "this one on the left").

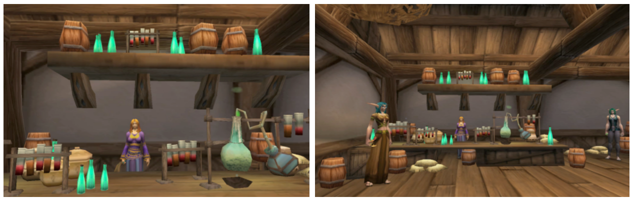

Figure 2. Near (left) and Far (right) object distances.

The third manipulation was Object Distance, which provided control for approximate visual distance to the objects of interest in the environment. Duplicate clips were recorded such that most objects available for reference in the environment were filmed at a perceptual distance of less than 9.9 WoW yards (Near) or between 11.11 and 28 WoW yards (Far). Figure 2 presents a still capture from two clips representing this manipulation.

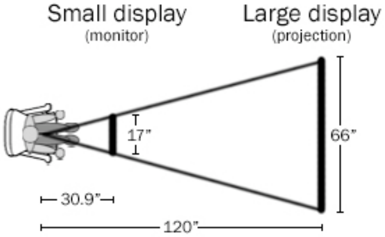

Figure 3. Configuration to maintain visual angle.

Apparatus

We used two displays, a 720p DLP projector and a 20" LCD monitor, in a physical setup similar to that described by Tan et al. [13]. Modifications are shown in Figure 3. The displays ran at wide-screen resolutions of perceptually equivalent brightness and contrast. The projected image was 66” wide by 40” tall, while the image on the monitor measured 17” by 10.5”. In order to maintain the same field of view across the displays, participants were seated 30.9” from the monitor and 120” away from the projection with the chair height adjusted so their eye-level was at the center of either display[1]. The experimenter sat immediately to the right of the participant within his or her field of view so that it was clear they could both see the display.

Measures

The spoken corpus contained a total of 24,123 words, with an average of 964 spoken words per participant. Two independent coders blind to experimental condition tagged all instances of local deixis (*this, these, here, here’s*) and remote deixis (*that, those, there, there’s*). Both references to objects and to the environment were tagged (e.g., “**This** is a very nicely tended garden area”), as well as uses of deixis following an object reference (“There are little trolls inside **there**”) and repeated references to objects (“**That** man–**that** dwarf…”). References to events or objects not present in the scene were not tagged (e.g., “We just ran into someone–that was cool”) nor were anaphoric uses of the terms (e.g., “She’s like an elf […] or something like that”). The coders overlapped in their ratings on 10% of the corpus and inter-rater reliability across the codes was high (κ=.86).

RESULTS

Initial examination of our qualitative data revealed multiple instances in which large display participants used local deixis and small display participants used remote deixis to describe the same visual content. The Description task excerpts below demonstrate this pattern. However, further investigation revealed a more nuanced distinction.

Large Display
So maybe **this** is the common room. Maybe they're making silk or something with **these** things.

Small Display
I can't remember what **those** things are called **there**…I think they're for cloth making?

Large Display
This looks like an area where they kill and torture people. **This** spinny vault […] is a bit ominous.

Small Display
The big circular thing above the gateway I guess–**there**. Not sure what **that** would–what **that’s** for.

Statistical Analysis

The analysis examines the rate of local and remote deixis generated as the primary dependent variables[2]. We performed a repeated measures analysis of variance in which Object Distance (Near or Far), Task (Description or ID) and Block (1-4) were repeated, and Display Size (Large or Small) was a between-participant factor. Word count, gender, age, visual-spatial ability, and MMOG experience were initially included as covariates along with all 2- and 3-way interactions. However, only word count and gender were found to be significant factors and all other covariates were dropped from the final models. As each participant completed four trials, observations were not independent of each other. Therefore, participant, nested within Display Size, was modeled as a random effect.

The model of the rate of *local* deixis achieved a good fit to the data (*Adj* R^2=.74, *p*<.001). Manipulation checks showed that the ID task produced a higher rate of local deixis than the Description task ($F_{(1,70)}$=85.3, *p*<.001). This was expected given that the ID task was designed to elicit object references. Similarly, the Object Distance manipulation showed a marginal increase in the rate of local deixis used for nearby stimuli ($F_{(1,70)}$=3.71, *p*=.058).

The Display Size manipulation did not reveal a main effect difference between the two display conditions ($F_{(1,18)}$=.61, n.s.). However, the effect was masked by a significant Display Size × Task interaction ($F_{(1,70)}$=16.51, *p*<.001; shown in Figure 4). Examination reveals that Display Size had considerable influence over the production of local deixis—but only when participants were engaged in the Description task. In this case, the rate of local deixis in the Large Display condition (*M*=2.25) was much greater than in the Small Display condition (*M*=.312; $F_{(1,28)}$=6.55, *p*=.016 for the contrast). However, when participants were engaged in the ID task there was no detectable difference in the rate of local deixis between Large (*M*=2.74) and Small Display conditions (*M*=3.62; $F_{(1,27)}$=1.36, *p*=.25 for the contrast).

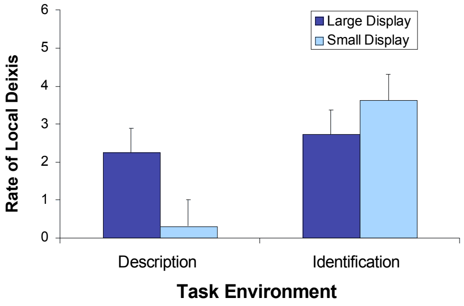

Figure 4. Local Deixis Rate by Display Size and Task.

Additionally, a significant Distance × Task interaction existed ($F_{(1,70)}$=4.22, *p*=.043). Participants engaged in the Description task produced a greater rate of local deixis when objects were near (*M*=1.98) than when they were far (*M*=.59; $F_{(1,70)}$=7.9, *p*=.006 for the contrast). However, in the ID task we found no difference in the rate of local deixis between Near (*M*=3.158) and Far conditions (*M*=3.21; $F_{(1,70)}$=.010, *p*=.92 for the contrast). Finally, gender had an effect on the rate of local deixis ($F_{(1,70)}$=19.75, *p*=.043) with females (*M*=1.46) producing fewer local references than males (*M*=3.0).

The model of the rate of *remote* deixis also achieved a reasonable fit to the data (*Adj* R^2=.43, *p*<.001). However, this was primarily driven by the control variable of word

[1] In order to facilitate naturalistic conversation, we followed precedent set by Tan and colleagues and did not fasten the participant’s head in place. However, we observed only small head movements throughout the study.

[2] While we present rate models in this paper, we also performed raw count models and found a nearly identical pattern of results, except that the Description task exhibited a significantly larger number of deictic references due to increased task length and amount of speech generated.

count, and marginally by the Task factor. Contrary to our expectations, we did not detect theoretically interesting differences between the two display conditions ($F_{(1,18)}$=.42, n.s.), or any higher-order interactions.

DISCUSSION

This study demonstrates that referential forms can change depending on something as subtle as the overall display size (controlling for visual angle). It is instructive to note that we did not find effects across all task types or perceptual distances. Rather, large display conditions were associated with greater rates of local deixis *only* when participants were engaged in a task involving open-ended narration. When participants were forced to refer to objects in a static array, this difference disappeared. These differences may be due to the referential constraints imposed by the task. In the Description task, participants spoke freely about objects and could describe them relative to themselves; in the ID task, they often had to describe objects relative to one another.

Moreover, while this pattern of results occurred for both rates and overall counts of local deixis, we found no evidence to suggest that display size influenced patterns of remote deixis. This is similar to Kramer et al.'s [8] findings, which show a stronger link between local deixis and presence than between remote deixis and presence. It may be that when participants feel a greater degree of presence, their language reflects a higher engagement with the task. Further investigation is needed to tease these ideas apart.

Our results suggest that the combination of display configuration and task attributes can yield a more immersive experience, or greater sense of presence, that translates into measurable differences in language use. It is therefore important to consider all of these factors when designing interactive visual environments. For example, if the goal is to develop conversational agents that interact naturally with humans, then agents need to not only understand referential patterns, but also generate natural responses and initial descriptions that match their speaking partner's situation model. Failing this, people may overcompensate in unnatural ways [9], making successful interaction difficult. The design of collaborative technology can also be informed by a better understanding of how visual context influences linguistic behavior, particularly in situations with high potential for referential ambiguity (e.g., remote surgery or collaborative physical tasks).

CONCLUSION

In this paper we show that subtle differences such as display size can influence language use, yet these linguistic adaptations are sensitive to particular task attributes. These findings are crucial when considering the design of collaborative interaction spaces and highlight the need to better understand the relationship between technological affordances and language. Future research could investigate the ways in which linguistic measures can be used as indices of cognitive mindset in interactive technologies.

ACKNOWLEDGMENTS

This research was funded by National Science Foundation grant #IIS-0705901 and Microsoft Research. We thank Chris Karr, Isaac Wilson, Josh Waytz, Alex Markov, and our reviewers for their valuable insights.

REFERENCES

[1] Byron, D.K. & Stoia, L. (2005). An analysis of proximity markers in collaborative dialog. In *Proc. of 41st annual meeting of the Chicago Linguistic Society*.

[2] Cassell, J. (2004). Towards a model of technology and literacy development: Story listening systems. *Journal of Applied Developmental Psychology*, *25(1)*, 75-105.

[3] Czerwinski, M., Smith, G., Regan, T., Myers, B., Robertson, G. & Starkweather, G. (2003). Toward characterizing the productivity benefits of very large displays. In *Proc. of Interact 2003*, 9-16.

[4] Dudfield, H.J., Macklin, C., Fearnley, R., Simpson, A. & Hall, P. (2001). Big is better? Human factors issues of large screen displays with military command teams. In *Proc. of People in Control*, 304-309.

[5] Ekstrom, R.B., French, J.W., Harman, H. & Dermen, D. (1976). Kit of factor-referenced cognitive tests, Educational Testing Service, Princeton, NJ.

[6] Fussell, S.R., Setlock, L.D., Yang, J., Ou, J., Mauer, E.M. & Kramer, A. (2004). Gestures over video streams to support remote collaboration on physical tasks. *Human-Computer Interaction*, *19*, 273-309.

[7] Guimbretière, F. (2002). Fluid interaction for high resolution wall-size displays. PhD Dissertation. Stanford University, Stanford, CA.

[8] Kramer, A., Oh, L.M. & Fussell, S.R. (2006). Using linguistic features to measure presence in computer-mediated communication. In *Proc. of CHI 2006*, 913-916. NY: ACM Press.

[9] Oviatt, S., Levow, G.-A., Moreton, E. & MacEachern, M. (1998). Modeling global and focal hyperarticulation during human-computer error resolution. *Journal of the Acoustical Society of America*, *104(5)*, 3080-3091.

[10] Ranjan, A., Birnholtz, J.P. & Balakrishnan, R. (2007). Dynamic shared visual spaces: Experimenting with automatic camera control in a remote repair task. In *Proc. of CHI 2007*, 1177-1186. NY: ACM Press.

[11] Ruddle, D.M., Payne, S. & Jones, D. (1999). The effects of maps on navigation and search strategies in very-large-scale virtual environments. *Journal of Experimental Psychology: Applied*, *5*, 54-75.

[12] Tan, D.S., Gergle, D., Scupelli, P. & Pausch, R. (2006). Physically large displays improve performance on spatial tasks. *ACM Transactions on Computer-Human Interaction*, *13(1)*, 71-99.

[13] Tan, D.S., Gergle, D., Scupelli, P. & Pausch, R. (2003). With similar visual angles, larger displays improve spatial performance. In *Proc. of CHI 2003*, 217-224.